Journal of Applied Informatics and Technology **2026**, 8(1), 255530

RESEARCH ARTICLE

Transfer Learning-Based Segmentation of Pneumonia from Chest X-Rays Images

Kyi Pyar[1,*]

[1]University of Computer Studies, Mandalay, Myanmar

*Corresponding author: Kyi Pyar, kyipyar@ucsm.edu.mm

Article Information

Article History:
Received: 26 January 2024
Revised: 3 May 2024
Accepted: 12 May 2024
Published: 19 October 2025

Keywords:
Deep learning
DeepLabV3
Pneumonia Segmentation
SegNet

Abstract

Pneumonia remains a significant global health concern, warranting precise and efficient diagnostic tools. This study introduces a comprehensive approach to pneumonia segmentation leveraging advanced deep learning techniques. The primary goal is to enhance the precision of pneumonia localization within medical images, specifically chest X-rays, through the utilization of state-of-the-art deep learning models. This study explores the application of advanced segmentation models, namely DeepLabV3 and SegNet, for the automated identification and delineation of pneumonia-affected regions within chest X-ray images. DeepLabV3, renowned for its semantic segmentation capabilities that partitions an image into multiple segments or regions, each of which is associated with a specific semantic label, and SegNet, featuring an encoder-decoder that consists of two main components: an encoder and a decoder, are selected as the segmentation models. The training process of the system leverages the widely acknowledged Kermany dataset, specifically composed of chest X-ray images depicting cases of pneumonia. This dataset is well-established and holds a prominent status within the field, recognized for its relevance and significance in the context of pneumonia detection and classification tasks. As per the evaluation findings, it is evident that the system attains enhanced accuracy by 0.844 and Intersection over Union score of 0.81 when employing the DeepLabV3 architecture compared to the SegNet architecture.

1. Introduction

Pneumonia, characterized by inflammation of the lung tissue, imposes a significant burden on healthcare systems globally. This medical condition involves the inflammation of lung tissue, specifically affecting the air sacs, or alveoli. Typically, this inflammation arises from infections, including viruses, bacteria, or fungi. Pneumonia manifests in diverse forms, varying in severity and presenting symptoms such as cough, chest pain, difficult breathing, and fever.

There are three main types of pneumonia: viral pneumonia, bacterial pneumonia, and fungal pneumonia. Viral pneumonia, exemplified by illnesses like COVID-19, primarily spreads through respiratory droplets, contact with contaminated surfaces, or close interaction with an infected individual. Bacterial pneumonia can spread through respiratory droplets, transmitted when an infected person coughs or sneezes. On the other hand, fungal pneumonia is often associated with environmental exposure and may not directly transmit between individuals. Diagnostic tools commonly employed in pneumonia cases include chest X-rays and computed tomography (CT) scans, providing visualizations of lung abnormalities. Additionally, blood tests play a crucial role in identifying the causative agent, be it bacteria or viruses. Sputum tests are conducted to analyze the material produced during coughing, aiding in the accurate diagnosis and targeted treatment of pneumonia.

The understanding of pneumonia's diverse forms and transmission methods is pivotal for effective healthcare management and preventive measures, especially in the context of widespread respiratory infections. Early and precise diagnosis, facilitated by a combination of imaging techniques and laboratory tests, is essential for

ISSN 3088-1803 |

implementing timely and appropriate interventions to alleviate the impact of pneumonia on individuals and healthcare systems. Chest X-rays serve as a foundational diagnostic tool for pneumonia, offering crucial insights into the presence and extent of pulmonary abnormalities. Despite their significance, the manual interpretation of these radiographic images is labor-intensive and introduces variability among different observers. This variability has prompted the exploration of innovative solutions to enhance diagnostic efficiency. Swift and accurate diagnosis are pivotal for effective treatment, but traditional methods, particularly manual interpretation of chest X-ray images, present challenges in terms of time, subjectivity, and resource demands.

Traditional diagnostic approaches, reliant on manual interpretation of chest X-rays, are not only time-consuming but also susceptible to interobserver variability. The introduction of deep learning provides a promising avenue to automate and optimize the pneumonia segmentation process, potentially revolutionizing the diagnosis of respiratory ailments. Deep learning architectures, renowned for their ability to discern intricate patterns in images, offer a promising solution to overcome the limitations associated with manual interpretation. In response to these challenges, this paper investigates the application of deep learning for pneumonia segmentation using chest X-ray images. The aim is to leverage the capabilities of deep learning models, such as SegNet and DeepLabV3, to autonomously and accurately identify pneumonia-affected regions. By doing so, the study seeks to contribute to the evolution of diagnostic methodologies, providing a more efficient and reliable approach to pneumonia detection in chest X-ray images. By harnessing the capabilities of these advanced neural network architectures, we aim to develop a robust and accurate segmentation system that can contribute to the automation of pneumonia diagnosis. The key contributions of our work can be summarized as follows:

- Transfer Learning Framework: We propose a transfer learning-based approach that harnesses the representational power of pre-trained CNNs, SegNet and DeepLabV3, for pneumonia segmentation. By initializing the networks with weights learned from large-scale natural image datasets, we exploit rich feature representations that facilitate effective learning of pneumonia-specific features from Chest X-rays images.

- Model Adaptation and Fine-Tuning: We meticulously adapt and fine-tune the SegNet and DeepLabV3 architectures to suit the task of pneumonia segmentation. Through iterative experimentation and parameter optimization, we tailor the networks to effectively capture subtle patterns and spatial dependencies characteristic of pneumonia regions in Chest X-rays images.

- Clinical Relevance and Impact: Our proposed framework holds significant clinical relevance by offering a reliable and efficient tool for automated pneumonia segmentation from Chest X-rays images. By expediting the diagnostic process and assisting radiologists in identifying pathological regions, our approach has the potential to enhance patient care, particularly in resource-constrained healthcare settings where access to expert radiological interpretation may be limited.

The section 2 critically reviews existing literature related to pneumonia segmentation, deep learning applications in medical imaging, and relevant methodologies. Section 3 delves into the theoretical underpinnings that form the foundation of the proposed pneumonia segmentation system. This includes a detailed exploration of the principles of deep learning, particularly focusing on the chosen models such as SegNet and DeepLabV3. Section 4 outlines the architecture and design of the proposed pneumonia segmentation system employing deep learning models. It details the methodology, data preprocessing steps, model selection, and the overall workflow of the system. Section 5 presents the results of the performance evaluation conducted on the proposed system. The final section, Section 6, concludes the paper by summarizing the key findings, discussing implications, and suggesting potential avenues for future research.

2. Related work

Zhang et al. (2021) centrally addresses the diagnosis of novel coronavirus pneumonia, utilizing CT images and implementing a deep-learning algorithm within a neural network for the precise detection and segmentation of the condition. The method employed contributes to a swift and accurate delineation of both lung structures and infected areas, thereby supporting medical professionals in diagnosing novel coronavirus pneumonia with increased efficiency. The approach presented in this study not only aids in accurate identification but also enhances screening processes, crucial for the effective management of pneumonia cases. To achieve effective segmentation of the infected area, the study introduces a novel ResAU-Net model, which builds upon the U-Net network structure. The incorporation of a pre-trained ResNet as an encoder enhances the model's feature extraction capabilities. Moreover, the introduction of an attention mechanism improves the network's focus on regions of interest, thereby reducing computational burdens and elevating prediction accuracy. Sub-pixel convolution is subsequently employed for the upsampling of the feature image, contributing to the refinement of the segmentation results. The study's results reveal a commendable prediction accuracy of 73.40% for the infected area in novel coronavirus pneumonia.

Wahyuningrum et al. (2023) introduced an innovative approach to lung segmentation in chest X-ray images, leveraging deep learning techniques with the FCA-Net (Fully Convolutional Attention Network) architecture. To address the intricate challenges associated with feature representation, attention modules—specifically spatial attention and channel attention—are seamlessly integrated into the Res2Net encoder. The experimentation phase utilizes a dataset of chest X-ray images obtained from Qatar University, accessible in the Kaggle repository. The dataset, consisting of 1500 chest X-ray images, each measuring 256 × 256 pixels, is partitioned into 10% for testing and 90% for training. The training data undergoes K-Fold Cross-validation with varying values of K from 2 to 10. The most promising results are achieved by employing a combination of spatial and channel attention in the K-Fold division with K set to 5. In this configuration, the Dice Similarity Coefficient (DSC) attains a notable value of 97.24%, indicating a high degree of similarity between the predicted and ground truth masks. Additionally, the Intersection over Union (IoU) reaches 94.66% in the testing data, signifying a robust performance in accurately delineating lung regions in chest X-ray images.

Gite et al. (2022) presented a novel approach with a comprehensive analysis and discussion focused on the implementation of U-Net in lung segmentation using X-ray images. The uniqueness of this work lies in its detailed comparison between U-Net and three other benchmark segmentation architectures, particularly in the context of diagnosing tuberculosis (TB) or other pulmonary lung diseases. The authors bring attention to the fact that many prior studies neglected segmentation before classification, potentially leading to data leakage issues. In instances where segmentation was incorporated before classification, U-Net emerged as the prevalent choice. However, this paper advocates for the efficacy of U-Net as a substitute, demonstrating superior accuracy and mean Intersection over Union (mean_iou) metrics. The results discussed in the paper highlight that U-Net achieved an accuracy of over 98% in lung segmentation, with a mean_iou of 0.95.

Rahman et al. (2022) presented a pioneering deep learning framework aimed at improving the precision of lung region segmentation in Chest X-Ray (CXR) images. The proposed methodology adopts a "divide and conquer" strategy, systematically breaking down original CXRs into smaller image patches. These patches undergo independent segmentation, and the resulting segmentations are aggregated to achieve a comprehensive segmentation of the entire lung region. The framework employs a two-model ensemble strategy to enhance segmentation accuracy. The first model utilizes a traditional Convolutional Neural Network (CNN) for classifying individual image patches. The classified patches are then merged to obtain a pre-segmentation. The second model incorporates a modified U-Net architecture, tailored specifically for segmenting individual patches. The outputs from this model contribute to another pre-segmented image. The fusion of these two pre-segmented images is executed through a binary disjunction operation, resulting in an initial segmentation. To refine the initial segmentation, the authors employ post-processing steps that involve traditional image processing techniques such as erosion, dilation, connected component labeling, and region-filling algorithms. These steps contribute to obtaining the final segmentation, ensuring increased accuracy and precision. Comprehensive evaluations of the proposed methodology were conducted using two publicly available datasets (MC, JPCL) and one proprietary dataset from The University of Texas Medical Branch (UTMB). These datasets encompass a diverse range of CXR images, demonstrating the versatility and effectiveness of the framework in achieving accurate lung region segmentation across varied image sources.

Zhao et al. (2021) introduced an automatic method designed for segmenting pulmonary parenchyma in chest CT images to aid in the analysis of texture features, ultimately assisting radiologists in diagnosing COVID-19. The proposed segmentation method seamlessly integrates a three-dimensional (3D) V-Net with a shape deformation module implemented using a spatial transform network (STN). The 3D V-Net is employed for end-to-end lung extraction, while the deformation module refines the V-Net output based on prior shape knowledge. The efficacy of this segmentation method is rigorously validated against manual annotations crafted by experienced operators. Experimental results showcase the remarkable performance of the proposed method, achieving a Dice similarity coefficient of 0.9796, sensitivity of 0.9840, specificity of 0.9954, and a mean surface distance error of 0.0318 mm when compared to the manual annotations. Moreover, the COVID-19 classification model, leveraging statistically significant radiomic features, demonstrates an impressive area under the curve (AUC) of 0.9470, sensitivity of 0.9500, and specificity of 0.9270.

Victor Ikechukwu et al. (2021) presented a comparative study focusing on the performance of pre-trained models, specifically VGG-19 and ResNet-50, in contrast to training a CNN from scratch. To address overfitting concerns, the study incorporates data augmentation and dropout regularization techniques. The analysis results indicate that appropriately fine-tuned pre-trained models achieved a recall rate of 92.03%, showcasing comparable performance to Iyke-Net, a CNN trained from scratch. This finding underscores the effectiveness of leveraging pre-trained models in medical image analysis tasks, particularly in scenarios where acquiring large labeled datasets and the computational resources necessary for training from scratch may present significant challenges.

Kayalibay et al. (2017) delved into the application of three-dimensional CNNs for medical image segmentation, with a specific focus on hand and brain MRI. In contrast to conventional CNNs that typically utilize two-dimensional kernels, recent advancements in medical image segmentation have emphasized the effectiveness of three-dimensional kernels. This approach enables a more comprehensive analysis of the inherent three-dimensional structure present in medical images. The study introduces a CNN-based method featuring three-dimensional filters applied to the segmentation of hand and brain MRI. To address specific challenges encountered in medical image segmentation, two modifications to an existing CNN architecture are proposed. The efficacy of the method is validated using data from both the central nervous system and the bones of the hand.

Cao and Zhao (2021) introduced an innovative segmentation algorithm based on the U-Net architecture, specifically designed to tackle the challenge of insufficient feature extraction when segmenting lungs in CXRs with opacities. The proposed method incorporates a Variational Autoencoder (VAE) into the convolutional layer, resulting in the development of the FVAE model. The FVAE model is capable of simultaneously capturing detailed local information and global context by integrating the features of the convolutional layer with those of the VAE. To further enhance the network's ability to accurately locate and recognize targets, the paper introduces a three-terminal attention mechanism. This mechanism incorporates both channel attention and a spatial attention mechanism modified by high-scale features. The integration of the three-terminal attention mechanism strengthens the model's performance, leading to improved segmentation accuracy. The proposed algorithm is rigorously tested on the SNIH and JSRT datasets, and the results demonstrate superior performance in terms of Accuracy, Recall, and F1-Score values when compared to other segmentation algorithms.

De Silva et al. (2022) presented an innovative multi-network ensemble method that incorporates a selector network, providing an advanced approach to lung segmentation. The selector network assumes a crucial role in evaluating segmentation outputs generated by multiple networks. On a patient-specific basis, the selector network intelligently chooses which segmentation outputs to fuse, ultimately forming the final segmentation for each patient. The candidate lung segmentation networks employed in this study consist of U-Net, incorporating five different encoder depths, and DeepLabV3+, featuring two distinct backbone networks (ResNet50 and ResNet18). The selector network itself is an image classifier built on the ResNet18 architecture. All training processes are conducted using the publicly available Shenzhen CR dataset, and the proposed ensemble method's performance is rigorously evaluated using two independent and publicly accessible CR datasets—Montgomery County (MC) and Japanese Society of Radiological Technology (JSRT). The results demonstrate that the Intersection-over-Union scores achieved by the proposed approach surpass the standard averaging ensemble method by 13% on MC and 5% on JSRT, highlighting the effectiveness of the multi-network ensemble method and the selector network in improving lung segmentation accuracy.

Fernandes et al. (2023) presented an innovative approach for the detection of pneumonic lungs from chest X-rays utilizing a CNN-based model. The model introduced in this study has the potential to significantly aid healthcare professionals in pneumonia diagnosis and treatment in real-world scenarios. The proposed method involves the development of a hybrid model combining EfficientNetB0 as a transfer learning-based model and a support vector machine (SVM) with hinge loss. Leveraging the feature extraction capabilities of the pre-trained EfficientNetB0 model, followed by classification using SVM, allows for accurate differentiation between abnormal and normal chest X-rays. Empirical results demonstrated the efficacy of the proposed model, with statistical findings indicating superior performance in terms of classification accuracy, precision, recall, and AUC values compared to existing state-of-the-art models. The achieved overall accuracy of 97% underscores the potential of the proposed approach for pneumonia detection and highlights its practical utility in clinical settings.

Dey (2022) introduced an innovative COV-XDCNN model, augmented with an external filter, designed to automate the diagnosis of diseases like COVID-19 and Viral Pneumonia. This model serves as a valuable tool to support healthcare workers, particularly during periods of outbreak, by swiftly and accurately analyzing chest radiography images. By leveraging advanced deep learning techniques, the proposed model aims to enhance diagnostic efficiency and accuracy in identifying respiratory illnesses. Empirical evaluation demonstrates the effectiveness of the proposed COV-XDCNN model with an external filter, achieving a remarkable test accuracy of 97.86% in classifying chest radiography images. Comparative analysis with other prominent models such as NASNetMobile, ResNet50, MobileNet, and VGG-16 underscores the superior performance of the proposed approach.

Kim et al. (2022) presented a novel fully automated framework aimed at aiding regional analysis through the utilization of deep learning-based four-region segmentation and detection models, specifically tailored for quantifying COVID-19 pneumonia. The framework employs a multi-step approach, beginning with a segmentation model to delineate the left and right lungs, followed by a detection network targeting the carina and left hilum for the separation of upper and lower lung regions. To enhance segmentation performance, an ensemble strategy integrating five models is employed. The clinical relevance of the proposed method is assessed by comparing it with the radiographic assessment of lung edema quality (RALE) as annotated by physicians. Evaluation metrics include mean intensities of the segmented four regions, which exhibit a positive correlation with the regional extent and density scores of pulmonary opacities derived from the RALE.

3. Background Theory

This section describes the theoretical background of this proposed system.

3.1 Segmentation

Segmentation, in the context of image processing and computer vision, refers to the process of dividing an image into meaningful and homogeneous regions or segments. The goal is to simplify or change the representation of an image into something more meaningful and easier to analyze. Image segmentation is a fundamental step in various computer vision applications, including object recognition, scene understanding, and medical image analysis. The primary objective of segmentation is to identify and delineate regions or objects within an image based on certain visual characteristics or properties. There are three types of segmentation:

- Semantic Segmentation: Assigns a label to each pixel in the image, providing a detailed understanding of the scene. It is commonly used in tasks like object recognition and scene understanding.
- Instance Segmentation: Identifies and distinguishes individual instances of objects in an image. It goes a step beyond semantic segmentation by providing a unique label for each object instance.
- Boundary or Edge Detection: Focuses on detecting boundaries between different regions or objects in an image.

There are three types of methods:

- Pixel-based Segmentation: Classifies each pixel individually based on its intensity, color, or texture features.
- Region-based Segmentation: Groups pixels into regions based on certain criteria, such as similarity in color or texture.
- Graph-based Segmentation: Treats pixels as nodes in a graph and uses graph algorithms to group pixels into segments.

3.2 SegNet

SegNet is a deep learning architecture designed specifically for semantic segmentation tasks in computer vision (Badrinarayanan et al., 2017). Developed by researchers at the University of Cambridge, SegNet is known for its efficiency in pixel-wise image segmentation. The primary goal of SegNet is to perform semantic segmentation, where the model assigns a class label to each pixel in an input image. This enables the detailed understanding of the scene and the identification of objects and their boundaries. SegNet follows an encoder-decoder architecture. The encoder is responsible for capturing high-level features from the input image, while the decoder reconstructs the segmented output from these features. The encoder employs a hierarchy of convolutional and pooling layers to capture features at different spatial resolutions. This hierarchical feature extraction is crucial for understanding both global and local context in the image.

SegNet uses skip connections between the encoder and decoder to retain detailed spatial information. These connections help in preserving fine-grained details that are often lost during downsampling operations. The encoder of SegNet is inspired by the VGG16 architecture, known for its simplicity and effectiveness in image classification tasks. The encoder is typically initialized with pre-trained weights from VGG16 on large-scale image classification datasets like ImageNet. This allows SegNet to benefit from the learned features. The architecture of SegNet is presented in Figure 1.

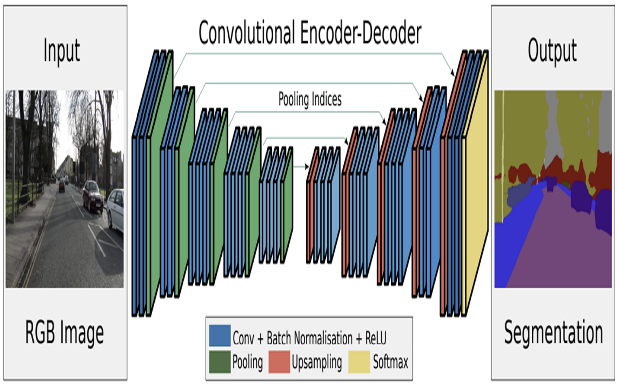


Figure 1. SegNet Architecture.

The decoder uses up-sampling layers to reconstruct the segmented output at the original resolution. These up-sampling layers are guided by the pooling indices obtained during the max-pooling operations in the encoder. The final layer of the decoder employs a softmax activation function to assign class probabilities to each pixel. The pixel is then labeled with the class having the highest probability. SegNet is trained using pixel-wise cross-entropy loss, which measures the difference between the predicted and ground truth pixel labels. This loss function is suitable for segmentation tasks where each pixel is assigned a class label. The model is trained using backpropagation and gradient descent optimization to minimize the loss function. In medical image analysis, SegNet is applied for the segmentation of organs, tumors, and other structures in various imaging modalities.

Advantages:

- Efficiency in Memory Usage: SegNet efficiently utilizes memory by employing a compact encoder-decoder architecture. Its encoder-decoder structure allows for efficient feature extraction and reconstruction, making it suitable for resource-constrained environments.

- Interpretability: SegNet's architecture facilitates interpretability due to its encoder-decoder design, where the encoder learns hierarchical features, and the decoder reconstructs the segmented image. This makes it easier for researchers and practitioners to understand the model's decision-making process.

- Suitability for Real-Time Applications: Due to its lightweight architecture and efficient inference process, SegNet is well-suited for real-time applications such as autonomous driving and video surveillance, where low-latency segmentation is essential.

Limitations:

- Limited Contextual Information: SegNet's symmetric encoder-decoder architecture may limit its ability to capture long-range contextual information, leading to potential loss of spatial details and context in the segmented output.

- Sensitivity to Input Variability: SegNet may be sensitive to input variability and noise, particularly in scenarios where there is significant variation in illumination, imaging conditions, or object scales. This sensitivity can affect the robustness and generalization capabilities of the model.

- Difficulty in Handling Class Imbalance: SegNet may struggle to handle class imbalance effectively, especially in datasets where certain classes are underrepresented. This can result in biased predictions and suboptimal performance, particularly for rare or minority classes.

3.3 DeepLabV3

DeepLabV3 is a state-of-the-art deep learning architecture designed for semantic image segmentation (Chen et al., 2017). Developed by Google, specifically by the Google Research Brain Team, DeepLabV3 builds upon its predecessors and incorporates advanced techniques to achieve highly accurate pixel-level segmentation. DeepLabV3 is primarily designed for semantic segmentation, a task where the model assigns a class label to each pixel in an input image. This enables a detailed understanding of the visual scene and identification of objects and their boundaries. DeepLabV3 is based on a deep CNN architecture that captures intricate features at different scales. The network leverages dilated convolutions and employs atrous spatial pyramid pooling (ASPP) to incorporate multi-scale contextual information. Dilated convolutions, also known as atrous convolutions, are utilized to increase the receptive field of the network without downsampling the spatial resolution. This helps the model capture contextual information at various scales. The architecture of DeepLabV3 is illustrated in Figure 2.

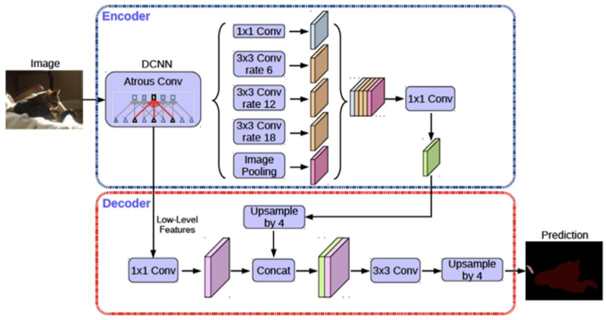


Figure 2. DeepLabV3 Architecture.

ASPP is a crucial component of DeepLabV3, allowing the model to gather information from multiple scales. It consists of parallel atrous convolutions with different dilation rates, enabling the network to aggregate features at different receptive fields. DeepLabV3 is flexible in terms of backbone networks. While it was initially designed with the MobileNetV2 backbone for efficiency, it can be adapted to use other powerful backbones such as ResNet or Xception for improved performance on high-resolution images. DeepLabV3 is trained using pixel-wise cross-entropy loss, which measures the dissimilarity between the predicted class probabilities and the ground truth labels for each pixel. Pre-training on large-scale datasets, such as ImageNet, is often employed to initialize the model with useful feature representations before fine-tuning on the segmentation task.

DeepLabV3 can benefit from post-processing techniques like CRF to refine segmentation results. CRF helps to smooth boundaries and improve the overall coherence of segmented regions. DeepLabV3 is widely used for scene understanding in computer vision applications. It can identify and segment objects in complex scenes, providing valuable information for higher-level tasks. DeepLabV3 achieves high accuracy in pixel-level segmentation, thanks to its multi-scale feature integration and context aggregation techniques. The architecture is flexible and can be adapted to different backbone networks, making it suitable for various image segmentation tasks.

Advantages:

- Multi-Scale Feature Integration: DeepLabV3 integrates multi-scale feature representations through dilated convolutions and atrous spatial pyramid pooling (ASPP), enabling the model to capture both local and global context effectively. This enhances its ability to segment objects of various sizes and scales.

- Semantic Segmentation Accuracy: DeepLabV3 achieves high semantic segmentation accuracy by leveraging advanced techniques such as dilated convolutions and atrous spatial pyramid pooling. This makes it suitable for applications where precise pixel-level segmentation is crucial, such as medical image analysis and autonomous driving.

- Flexibility and Adaptability: DeepLabV3 offers flexibility and adaptability through its modular design, allowing researchers to customize and fine-tune different components of the architecture according to specific task requirements and datasets. This flexibility facilitates experimentation and optimization for diverse segmentation tasks.

Limitations:

- Computational Complexity: DeepLabV3's advanced architectural features, such as dilated convolutions and ASPP, increase computational complexity, leading to higher inference times and resource requirements compared to simpler architectures like SegNet. This may limit its applicability in resource-constrained environments or real-time applications.

- Training Data Requirements: DeepLabV3 may require large amounts of labeled training data to effectively learn complex semantic features and achieve high segmentation accuracy. Obtaining and annotating large-scale datasets can be time-consuming and costly, particularly for medical imaging applications where expert annotations are required.

- Potential Overfitting: DeepLabV3's high capacity and complex architecture may be prone to overfitting, especially when trained on limited or noisy datasets. Regularization techniques and data augmentation strategies are necessary to mitigate the risk of overfitting and ensure generalization performance across diverse datasets.

4. Proposed Segmentation Architecture

The system presented in this study focuses on the development of an effective pneumonia segmentation system through the utilization of deep learning models, specifically SegNet and DeepLabV3.

The primary dataset employed is the Kermany dataset (Mooney, 2018), a collection of chest X-ray images sourced from the Guangzhou Women and Children's Medical Center. It is derived from chest radiographs obtained from pediatric patients undergoing diagnostic imaging procedures at the medical center. The dataset predominantly comprises chest X-ray images obtained from pediatric patients, including infants, children, and adolescents, presenting with various thoracic conditions and clinical indications. These images encompass a wide range of anatomical views, including frontal (posteroanterior, anteroposterior) and lateral projections, capturing different perspectives of the pediatric thoracic cavity. Each image may be associated with clinical metadata, such as patient demographics (age, gender), clinical history, presenting symptoms, and diagnostic findings, providing contextual information for image interpretation and analysis. The images in the dataset are typically labeled or annotated by expert pediatric radiologists or clinicians, providing ground truth information regarding the presence or absence of specific thoracic pathologies, abnormalities, or anatomical structures in pediatric patients. The dataset may encompass a spectrum of pediatric thoracic pathologies and conditions, including but not limited to pneumonia, bronchiolitis, congenital anomalies, respiratory distress syndrome, and other pediatric respiratory diseases. The dataset may be provided in standard image file formats (e.g., DICOM, PNG, JPEG) along with associated metadata and annotations, facilitating integration with medical image analysis workflows and software frameworks. The Kermany dataset, like many medical image datasets, likely contains images of various classes of chest X-rays, such as normal, pneumonia, pneumothorax, and other abnormalities. This dataset comprises a total of 5232 chest X-ray images, encompassing both normal and pneumonia-affected conditions.

The methodology initiates with the process of polyline annotation using LabelMe tools, ensuring precise identification and delineation of regions of interest within the X-ray images. Following this annotation, the json file annotations are converted into Mask images, creating a format suitable for segmentation tasks. Generate pixel-wise segmentation masks corresponding to each input image. In the case of the Kermany dataset, where segmentation is the task, these masks would indicate the regions of interest (e.g., lung boundaries, abnormalities) that the model needs to segment. The dataset is then strategically partitioned into two subsets – a training set constituting 75% of the data and a validation set comprising the remaining 25%. Several preprocessing steps are implemented to enhance the quality and suitability of the data for model training. These include resizing the images to a standardized 256×256 dimensions, and the crucial step of normalizing pixel intensity in both RGB images and Mask images. This normalization helps in reducing data variability and improving model convergence during training. These preprocessing steps are crucial for ensuring consistency and enhancing the suitability of the data for model training.

Subsequently, the preprocessed datasets are loaded, and the necessary transformations are applied to convert the data into tensor format, facilitating the subsequent training process. The segmentation model is trained using the deep learning models SegNet and DeepLabV3. The training process involves iteratively adjusting model parameters to optimize performance and achieve accurate segmentation of pneumonia-affected regions. The segmentation model, leveraging deep learning architectures like SegNet and DeepLabV3, undergoes training with the aim of achieving the desired accuracy. The proposed system employs a systematic and structured approach, as illustrated in Figure 3, to guide the sequential steps of the training process for pneumonia segmentation using deep learning models. The flowchart provides a visual representation of the key stages involved in the development and training of the segmentation model on chest X-ray images.

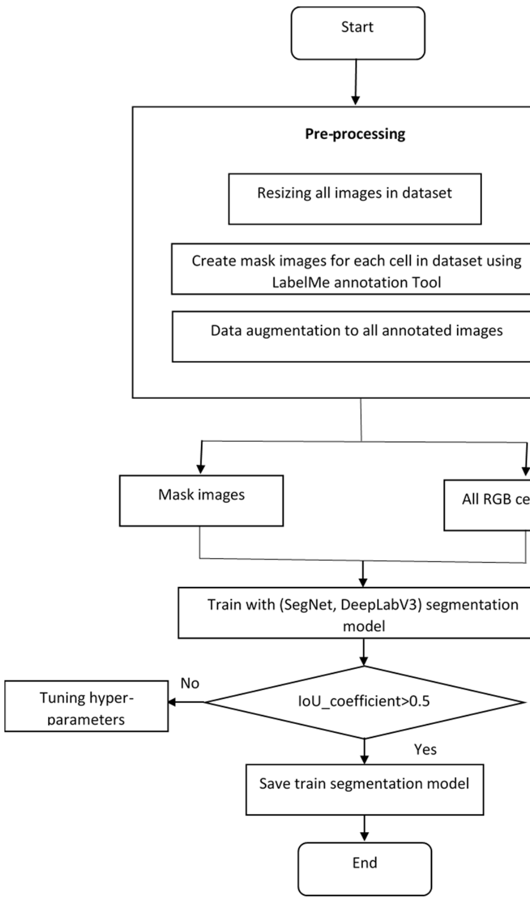


Figure 3. Flowchart of Training.

Throughout the training process, model parameters are iteratively adjusted to optimize performance, and the efficacy of the trained model is evaluated comprehensively. Various performance metrics, including IoU coefficient and accuracy, are employed to assess the model's ability to accurately segment pneumonia regions in chest X-ray images. Upon achieving the desired accuracy, the trained model is saved for future use, ensuring its applicability for subsequent tasks or deployments. This systematic and well-defined approach, as depicted in the flowchart, ensures the effectiveness of the deep learning models in pneumonia segmentation, providing a clear roadmap from data preparation to model training, evaluation, and preservation.

In the testing phase of the pneumonia segmentation system, the input image undergoes an initial resizing step as part of the preprocessing stage. This resizing operation ensures that the input image is brought to a standardized format suitable for further analysis. Subsequently, the trained segmentation model is loaded, leveraging the knowledge gained during the training phase, and the resized input image is subjected to the prediction process.

During prediction, the model evaluates the input image to identify regions indicative of pneumonia. If the model detects a pneumonia segment, it generates a segmented pneumonia cell binary image as the final output. This binary image serves as a visual representation of the identified pneumonia regions, with clear distinctions between affected and unaffected areas. The testing process is systematically outlined in Figure 4, providing a step-by-step illustration of the operations involved in resizing, prediction, and the generation of the segmented pneumonia cell binary image. This figure serves as a visual guide, emphasizing the key stages of the testing phase and the role each step plays in evaluating the model's performance on new or unseen data.

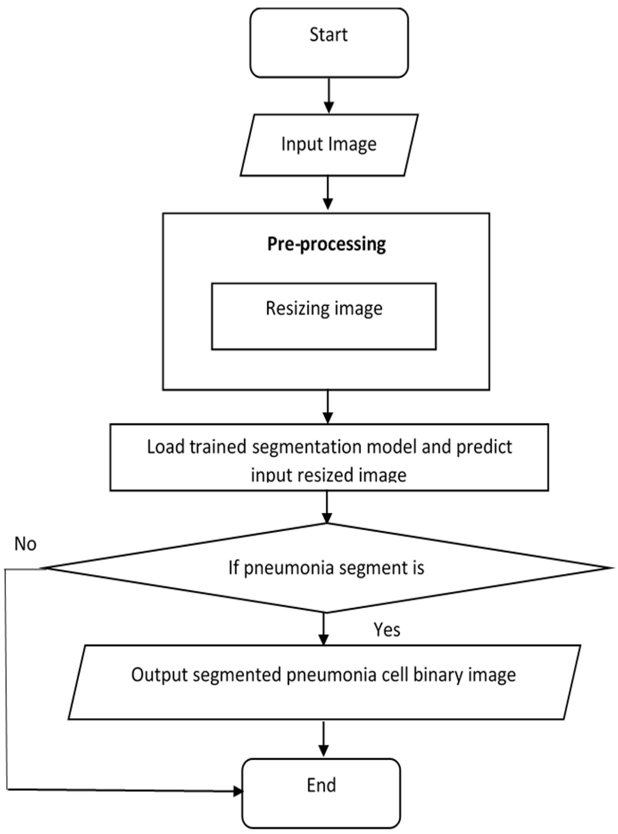


Figure 4. Flowchart of Testing.

The evaluation of the experiment involves a meticulous and comprehensive analysis, incorporating various indicators such as completeness, appropriateness, and quality. This assessment is designed to offer detailed insights into the performance metrics, with a specific emphasis on critical parameters including the IoU coefficient and accuracy. The primary objective of this evaluation is to measure the effectiveness of the deep learning segmentation model utilized in the study, providing a nuanced understanding of its performance across different criteria. Completeness, appropriateness, and quality serve as key benchmarks in the evaluation process, ensuring a thorough examination of the segmentation model's capabilities.

5. Experimental Results

In this system, the researchers employ the Chest X-Ray Images (Pneumonia) Kermany dataset as the foundational dataset, forming the backbone for training and testing the proposed system. This dataset comprises chest X-ray images categorized into two distinct groups: images depicting pneumonia and those illustrating normal (non-pneumonia) conditions. The dataset is strategically partitioned, with a ratio of 75% designated for training and 25% for testing. The partitioning process involves a random selection of 75% of the records for inclusion in the training dataset, leaving the remaining 25% for testing purposes. To thoroughly gauge the performance and efficacy of the proposed system, a rigorous evaluation is undertaken using the original dataset. This evaluation process includes the meticulous assessment of key performance metrics, notably the IoU coefficient and accuracy. The IoU coefficient serves as a crucial measure for evaluating the overlap between predicted and actual regions in image segmentation, providing insight into the segmentation model's accuracy. IoU measures the overlap between predicted and ground truth segmentation masks. By assessing the spatial alignment of segmented regions, IoU provides a precise indication of how well the model localizes objects of interest within the image. This is particularly critical in medical imaging, where accurate delineation of anatomical structures or pathological regions is essential for diagnosis and treatment planning. IoU is highly sensitive to segmentation quality, capturing even subtle deviations between predicted and ground truth masks. As a result, it offers a nuanced assessment of segmentation accuracy, allowing researchers and clinicians to identify areas where the model may be underperforming or producing inaccurate segmentations. IoU is mathematically represented by:

$$\text{IoU} = \frac{|A \cap B|}{|A \cup B|} \tag{1}$$

where, A represents the predicted segmentation mask, B represents the ground truth mask, $|A \cap B|$ denotes the area of intersection between the predicted and ground truth masks, and $|A \cup B|$ denotes the area of union of the predicted and ground truth masks.

Simultaneously, accuracy, a broader performance metric, is assessed to determine the overall correctness of the system's classifications. Accuracy is mathematically represented by:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{2}$$

where TP is the number of true positive pixels (correctly classified foreground pixels), TN is the number of true negative pixels (correctly classified background pixels), FP is the number of false positive pixels (foreground pixels incorrectly classified as background), and FN is the number of false negative pixels (background pixels incorrectly classified as foreground).

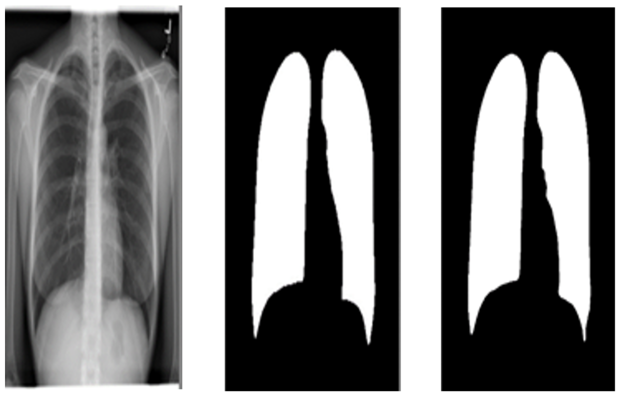
In the framework of this system, the segmentation models adopt distinct loss functions to optimize their training processes. For SegNet, Binary Cross Entropy Loss is employed—a widely used loss function particularly effective in binary classification tasks. This choice aligns with the model's focus on binary outcomes, enhancing its capability to differentiate between different classes, such as pneumonia and non-pneumonia regions in medical image segmentation. In contrast, DeepLabV3 utilizes the unet3p_hybrid_loss. This specific loss function is a hybrid combination that integrates Jaccard loss, SSIM loss (Structural Similarity Index), and Focal loss. The incorporation of these diverse elements in the loss function reflects a comprehensive strategy aimed at optimizing the training process for these segmentation models. Jaccard loss focuses on the intersection over union, SSIM loss evaluates structural similarity, and Focal loss addresses the issue of class imbalance by assigning higher weights to hard-to-classify examples.

Table 1. Parameters of SegNet and DeepLabV3.

Parameter	**SegNet**	**DeepLabV3**
Encoder	5 Convolutional layers with max-pooling	Backbone network: ResNet with atrous convolutions
Decoder	Mirrors encoder architecture with upsampling	Decoder
Activation function	ReLU	ReLU
Loss function	Binary Cross Entropy	unet3p_hybrid_loss
Optimizer	Adam	Adam
Learning rate	0.01	0.01
Atrous Spatial	None	Atrous Spatial Pyramid Pooling (ASPP)
Dropout	0.5	0.5
Batch size	8	8
Weight decay	0.0001	0.0001

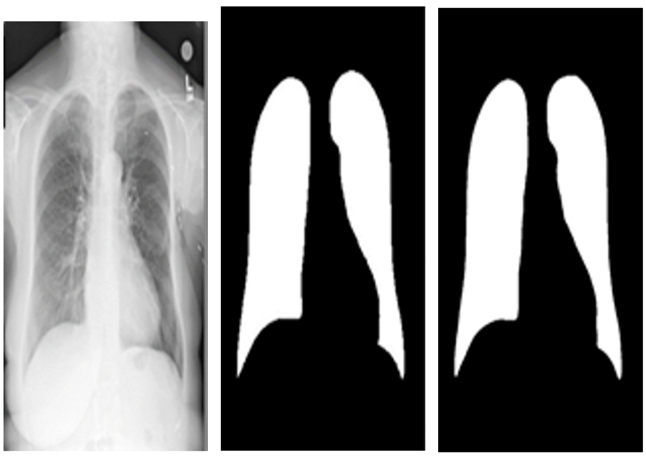
The selection of the unet3p_hybrid_loss underlines the intention to enhance overall segmentation performance by considering various aspects of model evaluation and optimization. Each loss function is tailored to meet the specific requirements and objectives of the corresponding segmentation model, contributing to a nuanced and effective training approach that encompasses multiple considerations for achieving accurate and robust segmentation results.

Table 1 describes the parameter of SegNet and DeepLabV3.



(a) (b) (c)

Figure 5. Segmentation using SegNet (a) Input Image (b) Predicted masks (c) Overlay masks.



(a) (b) (c)

Figure 6. Segmentation Using DeepLabV3 (a) Input Image (b) Predicted masks (c) Overlay masks.

Figure 5 presents the segmentation using SegNet. Figure 6 depicts the segmentation using DeepLabV3.

The comparative results for the performance of SegNet and DeepLabV3 are presented in Table 2.

Table 2. Performance results.

Architecture	Test Accuracy	Test IoU
DeepLabV3	0.844	0.81
SegNet	0.81	0.70

Figure 7 describes the accuracy results of SegNet and DeepLabV3. Figure 8 describes IoU results of SegNet and DeepLabV3. High IoU values signify strong agreement between predicted and ground truth segmentations, instilling confidence in the reliability and clinical relevance of the segmentation results. By leveraging IoU, researchers can gain valuable insights into the performance of their segmentation algorithms, driving advancements in medical image analysis and ultimately improving patient care.

According to the evaluation results, this performance of system using DeepLabV3 obtains the better accurate results than SegNet.

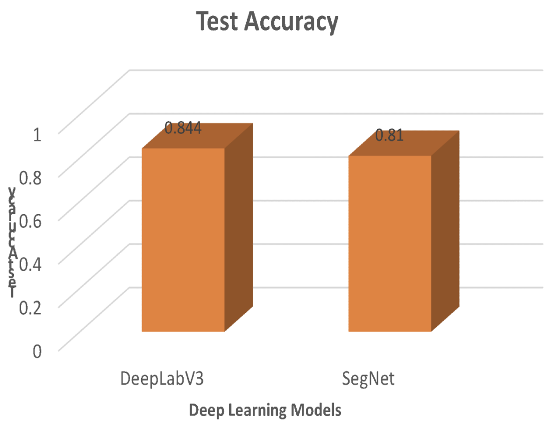


Figure 7. Accuracy Results.

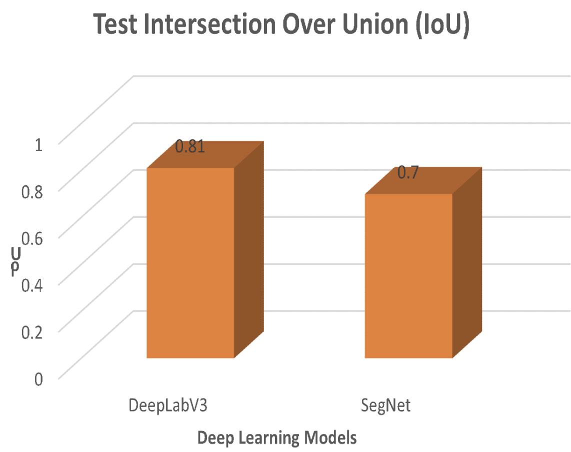


Figure 8. DeepLabV3 Architecture.

6. Conclusion

This research addresses the critical issue of pneumonia, a widespread respiratory infection, especially prevalent in underdeveloped and developing regions characterized by suboptimal living conditions and limited medical infrastructure. The study introduces a sophisticated approach to pneumonia classification by implementing a deep learning-based segmentation system employing SegNet, and DeepLabV3. The models are trained on the Chest X-Ray Images (Pneumonia) Kermany dataset. In the experimental segmentation phase, the study reports notable achievements in accuracy. Specifically, DeepLabV3 demonstrates an accuracy of 0.844, and SegNet achieves 0.81. Additionally, Intersection over Union (IoU) values are reported as 0.81 for DeepLabV3, and 0.70 for SegNet. These results emphasize that the system, particularly when employing DeepLabV3, surpasses the segmentation performance of SegNet in pneumonia detection. DeepLabV3 typically employs a deep convolutional neural network architecture, often with a backbone network: ResNet for feature extraction. It incorporates modules such as Atrous Spatial Pyramid Pooling (ASPP) to capture multi-scale contextual information. SegNet follows an encoder-decoder architecture, where the encoder extracts features from the input image, and the decoder reconstructs the segmented output. It typically consists of convolutional and pooling layers in the encoder and upsampling layers in the decoder. DeepLabV3 may incorporate advanced preprocessing techniques such as data augmentation, intensity normalization, and image resizing to enhance the quality and diversity of the training data, ensuring robustness and generalization capability. SegNet may apply simpler preprocessing steps such as normalization and resizing, focusing more on the model architecture for segmentation performance. However, the lack of advanced preprocessing may limit its ability to handle variations in image quality and appearance. Therefore, DeepLabV3 outperforms than SegNet in pneumonia segmentation. The study acknowledges the potential for further advancements and proposes future work focusing on enhancing the system's capabilities. A key aspect of this involves exploring the integration of an ensemble model, combining multiple deep learning approaches for pneumonia segmentation. Moreover, the segmentation for other datasets will be done with these deep learning methods.

References

Badrinarayanan, V., Kendall, A., and Cipolla, R. (2017). SegNet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(12):2481–2495. DOI: 10.1109/tpami.2016.2644615.

Cao, F. and Zhao, H. (2021). Automatic lung segmentation algorithm on chest X-ray images based on fusion variational auto-encoder and three-terminal attention mechanism. *Symmetry*, 13(5):814. DOI: 10.3390/sym13050814.

Chen, L., Papandreou, G., Schroff, F., and Adam, H. (2017). Rethinking atrous convolution for semantic image segmentation. *CoRR*, abs/1706.05587.

De Silva, M. S., Narayanan, B. N., and Hardie, R. C. (2022). A patient-specific algorithm for lung segmentation in chest radiographs. *AI*, 3(4):931–947. DOI: 10.3390/ai3040055.

Dey, A. (2022). *COV-XDCNN: Deep Learning Model with External Filter for Detecting COVID-19 on Chest X-rays*, page 174–189. Springer International Publishing. DOI: 10.1007/978-3-031-11633-9_14.

Fernandes, V., Bras, G., d, L. F., Paiva, e., Junior, G. B., d, A. C., Paiva, e., and Rivero, L. (2023). Transfer learning-based lung segmentation and pneumonia detection for paediatric chest X-ray images. *International Journal of Innovative Computing and Applications*, 14(1/2):56. DOI: 10.1504/ijica.2023.129358.

Gite, S., Mishra, A., and Kotecha, K. (2022). Enhanced lung image segmentation using deep learning. *Neural Computing and Applications*, 35(31):22839–22853. DOI: 10.1007/s00521-021-06719-8.

Kayalibay, B., Jensen, G., and van der Smagt, P. (2017). CNN-based segmentation of medical imaging data. *CoRR*, abs/1701.03056.

Kim, Y.-G., Kim, K., Wu, D., Ren, H., Tak, W. Y., Park, S. Y., Lee, Y. R., Kang, M. K., Park, J. G., Kim, B. S., Chung, W. J., Kalra, M. K., and Li, Q. (2022). Deep learning-based four-region lung segmentation in chest radiography for COVID-19 diagnosis. *Diagnostics*, 12(1):101. DOI: 10.3390/diagnostics12010101.

Mooney, P. (2018). Chest X-ray images (pneumonia). `https://www.kaggle.com/paultimothymooney/chest-xray-pneumonia`. Accessed: 2024-01-25.

Rahman, M. F., Zhuang, Y., Tseng, T.-L. B., Pokojovy, M., McCaffrey, P., Walser, E., Moen, S., and Vo, A. (2022). Improving lung region segmentation accuracy in chest X-ray images using a two-model deep learning ensemble approach. *Journal of Visual Communication and Image Representation*, 85:103521. DOI: 10.1016/j.jvcir.2022.103521.

Victor Ikechukwu, A., Murali, S., Deepu, R., and Shivamurthy, R. (2021). ResNet-50 vs VGG-19 vs training from scratch: A comparative analysis of the segmentation and classification of pneumonia from chest X-ray images. *Global Transitions Proceedings*, 2(2):375–381. DOI: 10.1016/j.gltp.2021.08.027.

Wahyuningrum, R. T., Yunita, I., Siradjuddin, I. A., Satoto, B. D., Sari, A. K., and Sensusiati, A. D. (2023). Improvement of chest X-ray image segmentation accuracy based on FCA-Net. *Cogent Engineering*, 10(1). DOI: 10.1080/23311916.2023.2229571.

Zhang, Z., Ni, X., Huo, G., Li, Q., and Qi, F. (2021). Novel coronavirus pneumonia detection and segmentation based on the deep-learning method. *Annals of Translational Medicine*, 9(11):934–934. DOI: 10.21037/atm-21-1156.

Zhao, C., Xu, Y., He, Z., Tang, J., Zhang, Y., Han, J., Shi, Y., and Zhou, W. (2021). Lung segmentation and automatic detection of COVID-19 using radiomic features from chest CT images. *Pattern Recognition*, 119:108071. DOI: 10.1016/j.patcog.2021.108071.